ABSTRACT OF THE DISCLOSURE

A text-to-speech system utilizes a method for producing a speech rendition of text based on dividing some or all words of a sentence into component diphones. A phonetic dictionary is aligned so that each letter within each word has a single corresponding phoneme. The aligned dictionary is analyzed to determine the most common phoneme representation of the letter in the context of a string of letters before and after it. The results for each letter are stored in a phoneme rule matrix. A diphone database is created using a wav editor to cut 2,000 distinct diphones out of specially selected words. A computer algorithm selects a phoneme for each letter. Then, two phonemes are used to create a diphone. Words are then read aloud by concatenating sounds from the diphone database. In one embodiment, diphones are used only when a word is not one of a list of pre-recorded words.